



IFWO

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/732,859

DATE: 08/27/2004

TIME: 14:29:10

Input Set : N:\Crf3\RULE60\10732859.raw
 Output Set: N:\CRF4\08272004\J732859.raw

1 <110> APPLICANT: J. Turck
 2 J. Archer
 3 <120> TITLE OF INVENTION: CONTROL OF GENE EXPRESSION IN EUKARYOTES
 4 <130> FILE REFERENCE: 9341-021
 5 <140> CURRENT APPLICATION NUMBER: US/10/732,859
 6 <141> CURRENT FILING DATE: 2003-12-09
 7 <150> PRIOR APPLICATION NUMBER: US/09/469,211
 8 <151> PRIOR FILING DATE: 1999-12-22
 9 <150> PRIOR APPLICATION NUMBER: UK 9828660.2
 10 <151> PRIOR FILING DATE: 1998-12-24
 11 <160> NUMBER OF SEQ ID NOS: 19
 12 <170> SOFTWARE: PatentIn Ver. 2.1
 13 <210> SEQ ID NO: 1
 15 <211> LENGTH: 7600
 16 <212> TYPE: DNA
 17 <213> ORGANISM: Rhodococcus sp.
 18 <220> FEATURE:
 19 <221> NAME/KEY: CDS
 20 <222> LOCATION: (295)..(1035)
 21 <223> OTHER INFORMATION: chpR regulator
 22 <221> NAME/KEY: CDS
 23 <222> LOCATION: (1261)..(2805)
 24 <223> OTHER INFORMATION: chpA transport
 W--> 25 <221> CDS
 26 <222> LOCATION: (2807)..(4720)
 27 <223> OTHER INFORMATION: chpB monooxygenase
 W--> 28 <221> CDS
 29 <222> LOCATION: (5721)..(6665)
 30 <223> OTHER INFORMATION: chpD catechol 2,3-dioxygenase
 W--> 31 <400> 1
 32 gaattccatg ttcttctcct tgcatgtggc ccgcgttggc gagggcactg gtcggcctg 60
 33 tcgcccccgag agggcgcatg tccgggtgcc tggatatggc gcgtacggcg tgccctccgg 120
 34 cgttaacccc gaggttggcc acgatgcccc gccatcagg tctggaatgc tagcgttcca 180
 35 gacgaaggta acccacagtg actcacacca caagtactag aatgcaagct gttgcgggtga 240
 36 ggcgcgcggc ataaggggga gccatgtccg ggacgcccac ggaaagcctg actcg 295
 37 atg acc acc acc gac acc ggc ccc aag ccg ggc agt gag gcc gcc 343
 38 Met Thr Thr Asp Thr Gly Pro Lys Pro Gly Ser Glu Ala Ala Ala
 39 1 5 10 15
 40 ctg ctc gcc aat gtc cgc acc tcg ggg gcg cgg ctg tcc tcc gcg ttg 391
 41 Leu Leu Ala Asn Val Arg Thr Ser Gly Ala Arg Leu Ser Ser Ala Leu
 42 20 25 30
 43 tac gac att ctg aag aac cgg ctg ctc gaa ggg cgc tat gcg gca ggc 439
 44 Tyr Asp Ile Leu Lys Asn Arg Leu Leu Glu Gly Arg Tyr Ala Ala Gly

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/732,859

DATE: 08/27/2004

TIME: 14:29:10

Input Set : N:\CrF3\RULE60\10732859.raw
 Output Set: N:\CRF4\08272004\J732859.raw

45	35	40	45	
46	gag aag atc gtc gtc gag tcg atc cgg caa gag ttc ggg gtg agc aag			487
47	Glu Lys Ile Val Val Glu Ser Ile Arg Gln Glu Phe Gly Val Ser Lys			
48	50	55	60	
49	cag ccc gtc atg gac gct ctg cgc cgc ctg tcc agc gac aag ctg gtc			535
50	Gln Pro Val Met Asp Ala Leu Arg Arg Leu Ser Ser Asp Lys Leu Val			
51	65	70	75	80
52	cac atc gtt ccc cag gtc ggt tgc gag gtc gtc tcc tac gcc ccg cgc			583
53	His Ile Val Pro Gln Val Gly Cys Glu Val Val Ser Tyr Ala Pro Arg			
54	85	90	95	
55	gaa gtg gaa gac ttc tac acc ctg ttc ggc ggt ttc gaa ggg acc atc			631
56	Glu Val Glu Asp Phe Tyr Thr Leu Phe Gly Gly Phe Glu Gly Thr Ile			
57	100	105	110	
58	gcc gcg gta gcg gcc tcc cgg cgg acc gag gcc cag ttg ctg gag ctg			679
59	Ala Ala Val Ala Ala Ser Arg Arg Thr Glu Ala Gln Leu Leu Glu Leu			
60	115	120	125	
61	gac ctg atc tcg gcg cgg gtc gac gcc ctg atc acc tcc cac gac ccg			727
62	Asp Leu Ile Ser Ala Arg Val Asp Ala Leu Ile Thr Ser His Asp Pro			
63	130	135	140	
64	gtg gtc cgc gcc cgc ggg tac cgc gtg cac aac cgg gag ttc cat gcg			775
65	Val Val Arg Ala Arg Gly Tyr Arg Val His Asn Arg Glu Phe His Ala			
66	145	150	155	160
67	gcc atc cac gcg atg gcg cac tcg cgg atc atg gag gag acc agc cag			823
68	Ala Ile His Ala Met Ala His Ser Arg Ile Met Glu Glu Thr Ser Gln			
69	165	170	175	
70	cga atg tgg gat ctg tcg gac ttc ttg atc aac acc acc ggc atc acc			871
71	Arg Met Trp Asp Leu Ser Asp Phe Leu Ile Asn Thr Thr Gly Ile Thr			
72	180	185	190	
73	aac ccg ctc tcg agc gca ctg ccc gac cgg cag cat gac cac cac gaa			919
74	Asn Pro Leu Ser Ser Ala Leu Pro Asp Arg Gln His Asp His His Glu			
75	195	200	205	
76	atc acc gag gcc atc cgc aac cgt gac gca gct gcc gcc cgc gag gcc			967
77	Ile Thr Glu Ala Ile Arg Asn Arg Asp Ala Ala Ala Arg Glu Ala			
78	210	215	220	
79	atg gaa cgc cac atc gtc ggc acc atc gca gta atc cgc gac gaa tcc			1015
80	Met Glu Arg His Ile Val Gly Thr Ile Ala Val Ile Arg Asp Glu Ser			
81	225	230	235	240
82	aac gcc cag ctg ccg agc tag accccgatac ccggggccatc gaccggctcc			1066
83	Asn Ala Gln Leu Pro Ser			
84	245			
85	gctatcgcgc cacctacgcc gagggggac tctcgccgt agcgctgcag acgatccacc			1126
86	ggcacccctcc acgctgaccc ctgtctgcc ctagaggggcc ggccgcgcgt cgatcacctt			1186
87	taccctcatc cagagacttg cgtcaccctc tatgcccag tagcgtctga actagacgtc			1246
88	tagcattcta gttga	gtg ctc cct ctc gaa gat tct cca gag		1288
89	Val Leu Pro Leu Glu Asp Ser Pro Glu			
90	250	255		
91	aac ccc tct cga aca tcc cca gaa gaa agg agc ggc cat gac gac cgc			1336
92	Asn Pro Ser Arg Thr Ser Pro Glu Glu Arg Ser Gly His Asp Asp Arg			
93	260	265	270	

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/732,859

DATE: 08/27/2004
TIME: 14:29:10

Input Set : N:\Crf3\RULE60\10732859.raw
Output Set: N:\CRF4\08272004\J732859.raw

94	ttc gca cgc atc gtc ctt cg ^g acg agc cca ctt cc ^g ccc aca gat	1384
95	Phe Ala Arg Ile Val Leu Arg Gly Thr Ser Pro Leu Pro Pro Thr Asp	
96	275 280 285	
97	cg ^g gga agc cc ^g acc gt ^g agc acc aca cct acc tcc cc ^g acg aag acc	1432
98	Arg Gly Ser Pro Thr Val Ser Thr Thr Pro Thr Ser Pro Thr Lys Thr	
99	290 295 300	
100	tca cc ^g ctg cg ^g gta gc ^g at ^g gc ^c ac ^t cc ^t gg ^t ac ^c cc ^c gtc gag	1480
101	Ser Pro Leu Arg Val Ala Met Ala Ser Phe Ile Gly Thr Thr Val Glu	
102	305 310 315	
103	ta ^c ta ^c ga ^c tt ^c at ^c ta ^c gg ^c ac ^c gc ^g gc ^g ct ^g gta tt ^c cc ^t	1528
104	Tyr Tyr Asp Phe Phe Ile Tyr Gly Thr Ala Ala Ala Leu Val Phe Pro	
105	320 325 330 335	
106	ga ^g tt ^g tt ^c tt ^c cc ^g ga ^t gtc tc ^g tcc gc ^g at ^c gg ^a at ^c ct ^g tt ^g tc ^g	1576
107	Glu Leu Phe Phe Pro Asp Val Ser Ser Ala Ile Gly Ile Leu Leu Ser	
108	340 345 350	
109	tt ^c gc ^g ac ^c tt ^c ag ^c gt ^t gg ^g tt ^c ct ^c gc ^c cc ^g ct ^g gg ^t gc ^c ata	1624
110	Phe Ala Thr Phe Ser Val Gly Phe Leu Ala Arg Pro Leu Gly Gly Ile	
111	355 360 365	
112	gt ^g tt ^c gg ^g ca ^c tt ^c gg ^t ga ^c cg ^g gtc gg ^c cc ^g aag ca ^g ca ^t ct ^g gt ^g	1672
113	Val Phe Gly His Phe Gly Asp Arg Val Gly Arg Lys Gln Met Leu Val	
114	370 375 380	
115	at ^c tcc ct ^g gtc gg ^a at ^g gg ^c tc ^g gc ^c ac ^c gt ^a ct ^g at ^g gg ^a tt ^g tt ^g	1720
116	Ile Ser Leu Val Gly Met Gly Ser Ala Thr Val Leu Met Gly Leu Leu	
117	385 390 395	
118	cc ^c gg ^t ta ^c gc ^c ca ^a at ^c gg ^g at ^c gc ^c cc ^c at ^c ct ^g ct ^g ac ^c ct ^g	1768
119	Pro Gly Tyr Ala Gln Ile Gly Ile Ala Ala Pro Ile Leu Leu Thr Leu	
120	400 405 410 415	
121	ct ^g cc ^c ct ^g gt ^g ca ^g gc ^c tt ^t gc ^c gg ^c gg ^c ga ^g tg ^g gg ^t gg ^a gc ^c	1816
122	Leu Arg Leu Val Gln Gly Phe Ala Val Gly Gly Glu Trp Gly Gly Ala	
123	420 425 430	
124	ac ^c ct ^g at ^g gc ^c gtc ga ^g ca ^c gc ^c cc ^c ac ^c gc ^c aag aa ^g gg ^c tt ^t tt ^c	1864
125	Thr Leu Met Ala Val Glu His Ala Pro Thr Ala Lys Lys Gly Phe Phe	
126	435 440 445	
127	gg ^a tcc tt ^c tcc ca ^g at ^g gg ^g ga ^c cc ^c gc ^c gg ^g ac ^c ag ^c gtc ga ^c acc	1912
128	Gly Ser Phe Ser Gln Met Gly Ala Pro Ala Gly Thr Ser Val Ala Thr	
129	450 455 460	
130	ct ^g gc ^c tt ^c gc ^c gtc tcc ca ^a tt ^g cc ^c ga ^c ga ^g ca ^g tt ^c ct ^g agt	1960
131	Leu Ala Phe Phe Ala Val Ser Gln Leu Pro Asp Glu Gln Phe Leu Ser	
132	465 470 475	
133	tgg gg ^c tgg ca ^g ct ^g cc ^g tt ^c ct ^g ac ^c gc ^c gt ^g ct ^g at ^c gt ^g at ^c	2008
134	Trp Gly Trp Arg Leu Pro Phe Leu Phe Ser Ala Val Leu Ile Val Ile	
135	480 485 490 495	
136	gg ^g ct ^g tt ^c att ca ^c ct ^g tcc ct ^g gc ^c ga ^a ag ^c cc ^c ga ^c tt ^c gc ^c ga ^g	2056
137	Gly Leu Phe Ile Arg Leu Ser Leu Ala Glu Ser Pro Asp Phe Ala Glu	
138	500 505 510	
139	gt ^g aag ga ^c ca ^g ag ^c gc ^c gt ^g gt ^g ca ^t cc ^g at ^c gc ^c ga ^a gc ^c tt ^c	2104
140	Val Lys Ala Gln Ser Ala Val Val Arg Met Pro Ile Ala Glu Ala Phe	
141	515 520 525	
142	cg ^c aag ca ^c tgg aag ga ^a att ct ^c ct ^c at ^c gc ^c acc tac ct ^g tcc	2152

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/732,859

DATE: 08/27/2004
TIME: 14:29:10

Input Set : N:\Crf3\RULE60\10732859.raw
Output Set: N:\CRF4\08272004\J732859.raw

143	Arg Lys His Trp Lys Glu Ile Leu Leu Ile Ala Gly Thr Tyr Leu Ser	
144	530 535 540	
145	caa gga gtg ttc gcc tat atc tgc atg gcc tac ctc gtc tcc tac ggc	2200
146	Gln Gly Val Phe Ala Tyr Ile Cys Met Ala Tyr Leu Val Ser Tyr Gly	
147	545 550 555	
148	acc acc gtc gcg ggg atc agc cgc acc ttc gcc ctg gcc gga gta ttc	2248
149	Thr Thr Val Ala Gly Ile Ser Arg Thr Phe Ala Leu Ala Gly Val Phe	
150	560 565 570 575	
151	gtc gcc ggc atc gtc gcc gtc ctc tac ctc gtg ttc ggc gct ctg	2296
152	Val Ala Gly Ile Val Ala Val Leu Leu Tyr Leu Val Phe Gly Ala Leu	
153	580 585 590	
154	tcc gac act ttc ggc cgc aag acc atg tac ctg ctc ggc gcc gcc gcg	2344
155	Ser Asp Thr Phe Gly Arg Lys Thr Met Tyr Leu Leu Gly Ala Ala Ala	
156	595 600 605	
157	atg ggt gtg gtg atc gcc ccc gcc ttc gca ctg atc aac acc ggc aac	2392
158	Met Gly Val Val Ile Ala Pro Ala Phe Ala Leu Ile Asn Thr Gly Asn	
159	610 615 620	
160	ccg tgg ctg ttc atg gcc gcg cag gtg ctg gtc ttc gga att gca atg	2440
161	Pro Trp Leu Phe Met Ala Ala Gln Val Leu Val Phe Gly Ile Ala Met	
162	625 630 635	
163	gcc ccc gcc ggc gtg aca ggc tcc ctg ttc acg atg gtc ttc gac	2488
164	Ala Pro Ala Ala Gly Val Thr Gly Ser Leu Phe Thr Met Val Phe Asp	
165	640 645 650 655	
166	gcf gac gtg cgc tac agc ggt gtc tct atc ggc tac acc atc tcc cag	2536
167	Ala Asp Val Arg Tyr Ser Gly Val Ser Ile Gly Tyr Thr Ile Ser Gln	
168	660 665 670	
169	gtc gcc ggc tcc gcg ttc gcc ccg acg atc gcg acc gcc ttg tac gcc	2584
170	Val Ala Gly Ser Ala Phe Ala Pro Thr Ile Ala Thr Ala Leu Tyr Ala	
171	675 680 685	
172	tcc acc aac acc agc aac tcg atc gtg acc tac ctg ctg atc gtc tcg	2632
173	Ser Thr Asn Thr Ser Asn Ser Ile Val Thr Tyr Leu Leu Ile Val Ser	
174	690 695 700	
175	gcc atc tcg atc gtc tcg gtg atc ctg ctg ccc ggc ggc tgg ggg cgc	2680
176	Ala Ile Ser Ile Val Ser Val Ile Leu Leu Pro Gly Gly Trp Gly Arg	
177	705 710 715	
178	aag ggc gct gcg agc cag ctc act cgc gac cag gcc acc tcc aca ccg	2728
179	Lys Gly Ala Ala Ser Gln Leu Thr Arg Asp Gln Ala Thr Ser Thr Pro	
180	720 725 730 735	
181	aaa atg cct gac acc gaa aca ttt tcg act cgg aca gtt ccg gac acc	2776
182	Lys Met Pro Asp Thr Glu Thr Phe Ser Thr Arg Thr Val Pro Asp Thr	
183	740 745 750	
184	gca gca tcc ctc gtc ctc gac aag tga a gtg atg aca gac atg agt	2825
185	Ala Ala Ser Leu Arg Val Leu Asp Lys Val Met Thr Asp Met Ser	
186	755 760 765	
187	gac cac gac cgc acc tcc tac gac acc gac gtc gtg atc gtc ggc ctc	2873
188	Asp His Asp Arg Thr Ser Tyr Asp Thr Asp Val Val Ile Val Gly Leu	
189	770 775 780	
190	ggc ccc gcc ggt ggc aca gcg gcg ctt gcc ctg gcc agc tac ggc atc	2921
191	Gly Pro Ala Gly Gly Thr Ala Ala Leu Ala Ser Tyr Gly Ile	

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/732,859

DATE: 08/27/2004
TIME: 14:29:10

Input Set : N:\Crf3\RULE60\10732859.raw
Output Set: N:\CRF4\08272004\J732859.raw

192	785	790	795	
193	cgc gtt cac gcc gtc tcg atg ttc ccc tgg gtg gcg aac tcg ccg cgc			2969
194	Arg Val His Ala Val Ser Met Phe Pro Trp Val Ala Asn Ser Pro Arg			
195	800	805	810	
196	gcg cac atc acc aac cag cgc gcc gtc gaa gtg ctg cgt gac ctg ggc			3017
197	Ala His Ile Thr Asn Gln Arg Ala Val Glu Val Leu Arg Asp Leu Gly			
198	815	820	825	830
199	gtc gaa gac gag gcg cgc aac tac gcc acc ccg tgg gac cag atg ggc			3065
200	Val Glu Asp Glu Ala Arg Asn Tyr Ala Thr Pro Trp Asp Gln Met Gly			
201	835	840	845	
202	gac acg ctg ttc acc acg agc ctg gcc ggc gag gag atc gtc cggt atg			3113
203	Asp Thr Leu Phe Thr Ser Leu Ala Gly Glu Glu Ile Val Arg Met			
204	850	855	860	
205	cag acc tgg ggt acg ggc gat atc cgc tac ggg gac tac ctg tcc gga			3161
206	Gln Thr Trp Gly Thr Gly Asp Ile Arg Tyr Gly Asp Tyr Leu Ser Gly			
207	865	870	875	
208	agc ccc tgc acg atg ctc gac att ccg cag ccc ctg atg gag ccg gtg			3209
209	Ser Pro Cys Thr Met Leu Asp Ile Pro Gln Pro Leu Met Glu Pro Val			
210	880	885	890	
211	ctg atc aag aac gcc gaa cgt ggt gcg gtc atc agc ttc aac acc			3257
212	Leu Ile Lys Asn Ala Ala Glu Arg Gly Ala Val Ile Ser Phe Asn Thr			
213	895	900	905	910
214	gaa tac ctc gac cac gcc cag gag gac ggg gtg acc gtc cggt ttc			3305
215	Glu Tyr Leu Asp His Ala Gln Asp Glu Asp Gly Val Thr Val Arg Phe			
216	915	920	925	
217	cgc gac gtc cgc tcg ggc acc gtg ttc acc cag cga gcc cgc ttc ctg			3353
218	Arg Asp Val Arg Ser Gly Thr Val Phe Thr Gln Arg Ala Arg Phe Leu			
219	930	935	940	
220	ctc ggt ttc gac ggc gca cga tcg aag atc gcc gaa cag atc ggg ctt			3401
221	Leu Gly Phe Asp Gly Ala Arg Ser Lys Ile Ala Glu Gln Ile Gly Leu			
222	945	950	955	
223	ccg ttc gaa ggt gaa ctc gcc cgc gcc ggt acc gcg tac atc ctg ttc			3449
224	Pro Phe Glu Gly Leu Ala Arg Ala Gly Thr Ala Tyr Ile Leu Phe			
225	960	965	970	
226	aac gcg gac ctg agc aaa tat gtc gct cat cgg ccg agc atc ttg cac			3497
227	Asn Ala Asp Leu Ser Lys Tyr Val Ala His Arg Pro Ser Ile Leu His			
228	975	980	985	990
229	tgg atc gtc aac tcg aag gcc ggt ttc ggt gag atc ggc atg ggt ctg			3545
230	Trp Ile Val Asn Ser Lys Ala Gly Phe Gly Glu Ile Gly Met Gly Leu			
231	995	1000	1005	
232	ctg cgc gcg atc cga ccg tgg gac cag tgg atc gcc ggc tgg ggc ttc			3593
233	Leu Arg Ala Ile Arg Pro Trp Asp Gln Trp Ile Ala Gly Trp Gly Phe			
234	1010	1015	1020	
235	gac atg gcg aac ggc gag ccg gat gtc tcc gac gac gtt gtc ctc gaa			3641
236	Asp Met Ala Asn Gly Glu Pro Asp Val Ser Asp Asp Val Val Leu Glu			
237	1025	1030	1035	
238	cag atc cgg acc ctc gtc ggc gac ccg cac ctg gac gtc gag atc gtg			3689
239	Gln Ile Arg Thr Leu Val Gly Asp Pro His Leu Asp Val Glu Ile Val			
240	1040	1045	1050	

VERIFICATION SUMMARY DATE: 08/27/2004
PATENT APPLICATION: US/10/732,859 TIME: 14:29:11

Input Set : N:\Crf3\RULE60\10732859.raw
Output Set: N:\CRF4\08272004\J732859.raw

L:25 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:28 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:31 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:689 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (7) SEQUENCE: